

AD-A260 069



12

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1373

July, 1992

Distance Metric between 3D Models and 2D Images  
for Recognition and Classification

Ronen Basri and Daphna Weinshall

DTIC  
ELECTE  
JAN 29 1993  
S C D

Abstract

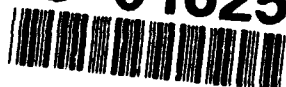
Similarity measurements between 3D objects and 2D images are useful for the tasks of object recognition and classification. We distinguish between two types of similarity metrics: metrics computed in image-space (*image metrics*) and metrics computed in transformation-space (*transformation metrics*). Existing methods typically use image metrics; namely, metrics that measure the difference in the image between the observed image and the nearest view of the object. Example for such a measure is the Euclidean distance between feature points in the image and their corresponding points in the nearest view. (Computing this measure is equivalent to solving the *exterior orientation calibration problem*.) In this paper we introduce a different type of metrics: *transformation metrics*. These metrics penalize for the deformations applied to the object to produce the observed image.

We present a *transformation metric* that optimally penalizes for "affine deformations" under weak-perspective. A closed-form solution, together with the nearest view according to this metric, are derived. The metric is shown to be equivalent to the Euclidean image metric, in the sense that they bound each other from both above and below. For the Euclidean image metric we offer a sub-optimal closed-form solution and an iterative scheme to compute the exact solution.

©Massachusetts Institute of Technology (1992)

This report describes research done at the Massachusetts Institute of Technology within the Artificial Intelligence Laboratory and the McDonnell-Pew Center for Cognitive Neuroscience. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-19-J-4038. Ronen Basri is supported by the McDonnell-Pew and the Rothchild post-doctoral fellowships. Daphna Weinshall is at IBM T.J. Watson Research Center, Hawthorne, NY.

93-01625



35P

DISTRIBUTION STATEMENT A

Approved for public release  
Distribution Unlimited

98 1 28 023

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE July 1992		3. REPORT TYPE AND DATES COVERED memorandum	
4. TITLE AND SUBTITLE Distance Metric Between 3D Models and 2D Images for Recognition and Classification				5. FUNDING NUMBERS N00014-85-K-0124 N00014-91-J-4038	
6. AUTHOR(S) Ronen Basri and Daphna Weinshall					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Artificial Intelligence Laboratory Massachusetts Institute of Technology 545 Technology Square Cambridge, Massachusetts 02139				8. PERFORMING ORGANIZATION REPORT NUMBER  AIM 1373	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of naval Research Information Systems Arlington, Virginia 22217				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES  None					
12a. DISTRIBUTION/AVAILABILITY STATEMENT  Distribution of this document is unlimited				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  Similarity measurements between 3D objects and 2D images are useful for the tasks of object recognition and classification. We distinguish between two types of similarity metrics: metrics computed in image-space ( <i>image metrics</i> ) and metrics computed in transformation-space ( <i>transformation metrics</i> ). Existing methods typically use image metrics; namely, metrics that measure the difference in the image between the observed image and the nearest view of the object. Example for such a measure is the Euclidean distance between feature points in the image and their corresponding points in the nearest view. (Computing this measure is equivalent to solving the <i>exterior orientation calibration problem</i> .) In this paper we introduce a different type of metrics: <i>transformation metrics</i> . These metrics penalize for the deformations applied to the object to produce the observed image.  (continued on back)					
14. SUBJECT TERMS				15. NUMBER OF PAGES 33	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UNCLASSIFIED		

Block 13 continued:

We present a *transformation metric* that optimally penalizes for "affine deformations" under weak-perspective. A closed-form solution, together with the nearest view according to this metric, are derived. The metric is shown to be equivalent to the Euclidean image metric, in the sense that they bound each other from both above and below. For the Euclidean image metric we offer a sub-optimal closed-form solution and an iterative scheme to compute the exact solution.

# 1 Introduction

Object recognition is a process of selecting the object model that best matches the observed image. A common approach to recognition uses features (such as points or edges) to represent objects. An object is recognized in this approach if there exists a viewpoint from which the model features coincide with the corresponding image features, e.g. [Roberts, 1965, Fischler and Bolles, 1981, Lowe, 1985, Huttenlocher and Ullman, 1987, Basri and Ullman, 1988, Thompson and Mundy, 1987, Ullman and Basri, 1991]. Since images often are noisy and models occasionally are imperfect, it is rarely the case that a model aligns perfectly with the image. Systems therefore look for a model that "reasonably" aligns with the image. Consequently, measures that assess the quality of a match become necessary. Such measures are useful also for the initial classification of novel objects, by associating the new object to the most similar object in the library.

A common measure for comparing 3D objects to 2D images is the Euclidean distance between feature points in the actual image and their corresponding points in the nearest view of the object. The assumption underlying this measure is that images are significantly less reliable than models, and so perturbations should be measured in the image plane. This assumption often suits recognition tasks. Other measures may better suit different assumptions. For example, when classifying objects, there is an inherent uncertainty in the structure of the classified object. One may therefore attempt to minimize the amount of deformations applied to the object to account for this uncertainty. Such a distance is measured in transformation space rather than in image space. A definition of these two types of measures is given in Section 3.

Measures to compare 3D models and 2D images generally are desired to have metrical properties; that is, they should monotonically increase with the difference between the measured entities. (A more exact definition is given in Appendix A.) The Euclidean distance between the image and the nearest view defines a metric. (We refer to this measure as the *image metric*.) The difficulty with employing this measure is that a closed-form solution to the problem has not yet been found, and therefore currently numerical methods must be employed to compute the measure. A common method to achieve a closed-form metric is to extend the set of transformations that objects are allowed to undergo from the rigid to the affine one. The problem with this measure is that it bounds the rigid measure from below, but not from above. Other methods either achieve only sub-optimal distances, or they do not define a metric. The existing approaches are reviewed in Section 2.

This paper presents a closed-form distance metric to compare 3D models and 2D images. The metric penalizes for the non-rigidities induced by the optimal affine transformation that aligns the model to the image under weak-perspective projection. The metric is shown to bound the least-square distance between the model and the image both from above and below. We foresee three ways to use the metric developed in this paper:

1. Obtain a direct assessment of the similarity between 3D models and 2D images.

Accession For	
NTIS <input checked="" type="checkbox"/> <del>Special</del>	
VIA <input checked="" type="checkbox"/> <del>Mail</del>	
Announced	
Justification	
By	
Distribution/	
Availability Code	
Dist	Avail and/or Special
A-1	

2. Obtain lower and upper bounds on the *image metric*. In many cases such bounds may suffice to unequivocally determine the identity of the observed object.
3. Provide an initial guess to be then used by a numerical procedure to solve the image distance.

The rest of this paper is organized as follows: In Section 2 we review related work. In Section 3 we define the concepts used in this paper. In Section 4 we summarize the main results of this paper. These results are discussed in detail and proved in section 5 for the *transformation metric*, and section 6 for the *image metric*. Sections 5 and 6 can be omitted in first reading. Finally, in Section 7 we compare the distances between 3D objects and 2D images, obtained by alignment, to our results.

## 2 Previous approaches

Previous approaches to the problem of model and image comparison using point features are divided into three major categories:

1. Least-square minimization in image space.
2. Sub-optimal methods using correspondence subsets.
3. Invariant functions.

The traditional photometric approach to the problem of model and image comparison involves retrieving a view of the object that minimizes the least-square distance to the image. This problem is referred to as the *exterior orientation calibration problem* (or the recovery of the *hand-eye transform*) and is defined as follows. Given a set of  $n$  3D points (model points) and a corresponding set of  $n$  2D points (image points), find the rigid transformation that minimizes the distance in the image plane between the transformed model points and the image points. An analytic solution to this problem has not yet been found. (Analytic solutions to the *absolute orientation problem*, the least-square distance between pairs of 3D objects, have been found, see [Horn, 1987, Horn, 1991]. An analytic solution to the least-square distance between pairs of 2D images has not yet been found.) Consequently, numerical methods are employed (see reviews in [Tsai, 1987, Yuan, 1989]). Such solutions often suffer from stability problems, they are computationally intensive and require a good initial guess.

To avoid using numerical methods, frequently the object is allowed to undergo affine transformations instead of just rigid ones. Affine transformations are composed of general linear transformations (rather than rotations) and translations, and they include in addition to the rigid transformations also reflection, stretch, and shear. The solution in the affine case is simpler than that of the rigid case because the quadratic constraints imposed in the rigid case are not taken into account, enabling the construction of a closed-form solution. At least six points

are required to find an affine solution under perspective projection [Fischler and Bolles, 1981], and four are required under orthographic projection [Ullman and Basri, 1991].

The affine measure bounds the rigid measure from below. The rigid measure, however, is not bounded from above, as is demonstrated by the following example. Consider the case of matching four model points to four image points under weak-perspective. Since in this case there always exists a unique affine solution, the affine distance between the model and the image is zero. On the other hand, since three points uniquely determine the rigid transformation that aligns the model to the image, by perturbing one point we can increase the rigid distance unboundedly.

A second approach to comparing models to images involves the selection of a small subset of correspondences (*alignment key*), solving for the transformation using this subset, and then transforming the other points and measuring their distance from the corresponding image points. Three [Fischler and Bolles, 1981, Rives et al., 1981, Haralick et al., 1991] or four [Horaud et al., 1989] points are required under perspective projection, and three points under weak perspective [Ullman, 1989, Huttenlocher and Ullman, 1987]. The obtained distance critically depends on the choice of alignment key. Different choices produce different distance measures between the model and the image. The results almost always are sub-optimal, since it is generally better to match all points with small errors than to exactly match a subset of points and project all the errors onto the others.

A third approach involves the application of invariant functions. Such functions return a constant value when applied to any image of a particular model. Invariant functions were successfully used only with special kinds of models, such as planar objects (e.g., [Lamdan et al., 1987, Forsyth et al., 1991]). More general objects can be recognized using model-based invariant functions [Weinshall, 1993]. For noise-free data, model-based invariant functions return zero if the image is an exact instance of the object. To account for noise, the output of these functions usually is required to be below some fixed threshold. In general, very little research has been conducted to characterize the behavior of these functions when the model and the image do not perfectly align. The result of thresholding therefore becomes arbitrary.

### 3 Definitions and notation

In the following discussion, we assume weak-perspective projection. Namely, the object undergoes a 3D transformation that includes rotation, translation, and scale, and is then orthographically projected onto the image. Perspective distortions are not accounted for and treated as noise.

In order to define a similarity measure for comparing 3D objects to 2D images, as discussed in section 1, we first define the **best-view** of a 3D object given a 2D image:

**Definition 1:** [*best-view*] Let  $\partial$  denote a difference measure between two 2D images of  $n$  features. Given a 2D image of an object composed of  $n$  features, the **best-view** of a 3D object

(model) composed of  $n$  corresponding features, is the view for which the smallest value of  $\partial$  is obtained. The minimization is performed over all the possible views of the model; the views are obtained by applying a transformation  $T$ , taken from the set of permitted transformations  $\Lambda$ , and followed by a projection,  $\Pi$ .

We compute  $\partial$ , the difference between two 2D images of  $n$  features in two ways:

**image metric:** we measure position differences in the image, namely, it is the Euclidean distance between corresponding points in the two images, summed over all points.

**transformation metric:** the images are considered to be instances of a single 3D object. The metric measures the difference between the two transformations that align the object with the two images. This difference can be measured, for instance, by computing the Euclidean distance between the matrices that represent the two transformations (when the two transformations are linear).

As mentioned above, the measure  $\partial$  is applied to the given image and to the views of the given model. These views are generated by applying a transformation from a set  $\Lambda$  of permitted transformations. The view that minimizes the distance  $\partial$  to the image is considered as the best view, and the distance between the best view and the actual image is considered as the distance between the object and the image.

We consider in this paper two families of transformations: rigid transformations and affine transformations, and we discuss the following metrics:

$N_{im}$ : a metric that measures the image distance between the given image and the best rigid view of the object.

$N_{af}$ : a metric that measures the image distance between the given image and the best affine view of the object.

$N_{tr}$ : a transformation metric. We assume that the image is an affine view of the object. (When it is not, we substitute the image by the best affine view.) We look for the rigid view of the object so as to minimize the difference between the two transformations: the affine transformation (between the object and the image) and the rigid transformation (between the object and its possible rigid views.) In other words, we look for a view so as to minimize the amount of "affine deformations" applied to the object.

To illustrate the difference between image metrics and transformation metrics, Figure 1 shows an example of three 2D images, whose similarity relations reverse, depending on which kind of metric is used. Consider the planar object in Figure 1(b) as a reference object, and assume  $\Lambda$  contains the set of rigid transformations in 2D. The images in (a) and (c) are obtained by stretching the object horizontally (by 9/7) and vertically (by 3/2) respectively. (The image in (b) is obtained by applying a unit matrix to the object.)

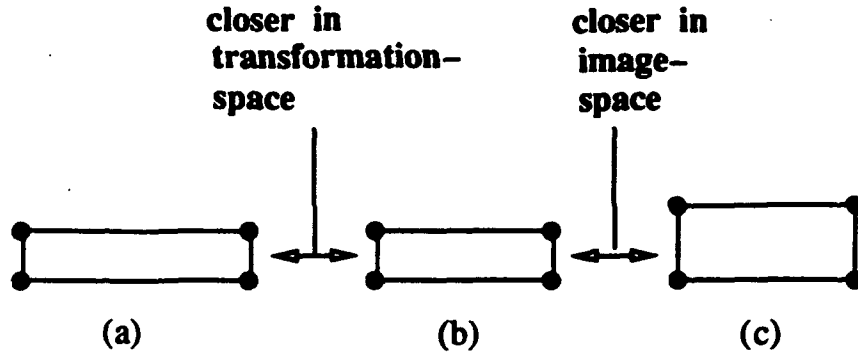


Figure 1: The 2D image shown in (b) is closer to the image in (a) when the difference is computed in transformation space, and closer to the image in (c) when the difference is the Euclidean difference between the two images.

- The *image metric* between the images in (b) and (a) is 4, two pixel at each of the left corners of the rectangle.  
The *image metric* between the images in (b) and (c) is 2, one pixel at each of the upper corners of the rectangle.  
Therefore, according to the *image metric*, Figure (c) is closer to (b) than (a) is.
- To compute the *transformation metric* consider the planar object illustrated in (b). We compute the difference between the matrices that represent the affine transformation from (b) to both (a) and (c) and the matrix that represent the best rigid transformation (in this case it is the unit matrix): (a) is obtained from (b) by a horizontal stretch of  $9/7$ . The *transformation metric* between (a) and (b) is therefore  $2/7 = 9/7 - 1$ .  
(c) is obtained from (b) by a vertical stretch of  $3/2$ . The *transformation metric* in this case is  $1/2 = 3/2 - 1$ .  
Therefore, according to the *transformation metric*, Figure (a) is closer to (b) than (c) is.

It is interesting to note that in this example the solution obtained by minimizing the *transformation metric* seems to better correlate with human perception than the solution obtained by minimizing the *image metric*.

### 3.1 Derivation of $N_{im}$ and $N_{af}$

We now define the rigid and the affine *image metrics* explicitly. Under weak-perspective projection, the position in the image,  $\bar{q}_i = (x_i, y_i)$ , of a model point  $\bar{p}_i = (X_i, Y_i, Z_i)$  following a rigid transformation is given by

$$q_i = \Pi(R\bar{p}_i + \bar{t})$$



where  $R$  is a scaled,  $3 \times 3$  rotation matrix,  $\vec{t}$  is a translation vector, and  $\Pi$  represents an orthographic projection. More explicitly, denote by  $\vec{r}_1^T$  and  $\vec{r}_2^T$  the top two row vectors of  $R$ , and denote  $\vec{t} = (t_x, t_y, t_z)$ ; we have that

$$\begin{aligned} x_i &= \vec{r}_1^T \cdot \vec{p}_i + t_x \\ y_i &= \vec{r}_2^T \cdot \vec{p}_i + t_y \end{aligned} \quad (1)$$

where

$$\begin{aligned} \vec{r}_1^T \cdot \vec{r}_2 &= 0 \\ \vec{r}_1^T \cdot \vec{r}_1 &= \vec{r}_2^T \cdot \vec{r}_2 \end{aligned} \quad (2)$$

The rigid metric,  $N_{im}$ , minimizes the difference between the two sides of Eq. (1) subject to the constraints (2).

When the object is allowed to undergo affine transformations, the rotation matrix  $R$  is replaced by a general  $3 \times 3$  linear matrix (denoted by  $A$ ) and the constraints (2) are ignored. That is

$$q_i = \Pi(A\vec{p}_i + \vec{t})$$

Denote by  $\vec{a}_1^T$  and  $\vec{a}_2^T$  the top two row vectors of  $A$ , we obtain

$$\begin{aligned} x_i &= \vec{a}_1^T \cdot \vec{p}_i + t_x \\ y_i &= \vec{a}_2^T \cdot \vec{p}_i + t_y \end{aligned} \quad (3)$$

The affine metric,  $N_{af}$ , minimizes the difference between the two sides of Eq. (3).

To define the rigid and the affine metrics, we first note that the translation component of both the best rigid and affine transformations can be ignored if the centroids of both model and image points are moved to the origin. In other words, we begin by translating the model and image points so that

$$\sum_{i=1}^n \vec{p}_i = \sum_{i=1}^n \vec{q}_i = 0 \quad (4)$$

We claim that now  $\vec{t} = 0$ . The proof is given in Appendix C.

Denote

$$P = \begin{pmatrix} X_1 & Y_1 & Z_1 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n \end{pmatrix}$$

a matrix of model point coordinates, and denote

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad (5)$$

the location vectors of the corresponding image points. A rigid metric that reflects the desired minimization is given by

$$N_{im} = \min_{\vec{r}_1, \vec{r}_2 \in \mathcal{R}^3} \|\vec{x} - P\vec{r}_1\|^2 + \|\vec{y} - P\vec{r}_2\|^2 \quad \text{s.t.} \quad \vec{r}_1^T \cdot \vec{r}_2 = 0, \quad \vec{r}_1^T \cdot \vec{r}_1 = \vec{r}_2^T \cdot \vec{r}_2 \quad (6)$$

The corresponding affine metric is given by

$$N_{af} = \min_{\vec{a}_1, \vec{a}_2 \in \mathcal{R}^3} \|\vec{x} - P\vec{a}_1\|^2 + \|\vec{y} - P\vec{a}_2\|^2 \quad (7)$$

In the affine case the solution is simple. We assume that the rank of  $P$  is 3 (the case for general, not coplanar, 3D objects). Denote  $P^+ = (P^T P)^{-1} P^T$ , the pseudo-inverse of  $P$ ; we obtain that

$$\begin{aligned} \vec{a}_1 &= P^+ \vec{x} \\ \vec{a}_2 &= P^+ \vec{y} \end{aligned} \quad (8)$$

And the affine distance is given by

$$N_{af} = \|(I - PP^+) \vec{x}\|^2 + \|(I - PP^+) \vec{y}\|^2 \quad (9)$$

Since the solution in the rigid case is significantly more difficult than the solution in the affine case, often the affine solution is considered, and the rigidity constraints are used only for verification (e.g. [Ullman and Basri, 1991, Weinshall, 1993, DeMenthon and Davis, 1992]).

The constraints (2) (substituting  $\vec{a}_i$  for  $\vec{r}_i$ , and using Eq. (8)) can be rewritten as

$$\begin{aligned} \vec{x}^T (P^+)^T P^+ \vec{y} &= 0 \\ \vec{x}^T (P^+)^T P^+ \vec{x} &= \vec{y}^T (P^+)^T P^+ \vec{y} \end{aligned}$$

Denote

$$B = (P^+)^T P^+ \quad (10)$$

we obtain that

$$\begin{aligned} \vec{x}^T B \vec{y} &= 0 \\ \vec{x}^T B \vec{x} &= \vec{y}^T B \vec{y} \end{aligned} \quad (11)$$

where  $B$  is an  $n \times n$  symmetric, positive-semidefinite matrix of rank 3. (The rank would be smaller if the object points are coplanar.)

We call  $B$  the **characteristic matrix** of the object.  $B$  is a natural extension to the  $3 \times 3$  model-based invariant matrix defined in [Weinshall, 1993]. A more general definition, and its efficient computation from images, is discussed in Appendix B.

### 3.2 Derivation of $N_{tr}$

We can now define a *transformation metric* as follows. Consider the affine solution. The nearest "affine view" of the object is obtained by applying the model matrix,  $P$ , to a pair of vectors,  $\bar{a}_1$  and  $\bar{a}_2$ , defined in Eq. (8). In general, this solution is not rigid, and so the rigid constraints (2) do not hold for these vectors. The metric described here is based on the following rule. We are looking for another pair of vectors,  $\bar{r}_1$  and  $\bar{r}_2$ , which satisfy the rigid constraints, and minimize the Euclidean distance to the affine vectors  $\bar{a}_1$ , and  $\bar{a}_2$ .  $P\bar{r}_1$  and  $P\bar{r}_2$  define the best rigid view of the object under the defined metric. The metric,  $N_{tr}$ , is defined by

$$N_{tr} = \min_{\bar{r}_1, \bar{r}_2 \in \mathcal{R}^3} \|\bar{a}_1 - \bar{r}_1\|^2 + \|\bar{a}_2 - \bar{r}_2\|^2 \quad \text{s.t.} \quad \bar{r}_1^T \cdot \bar{r}_2 = 0, \quad \bar{r}_1^T \cdot \bar{r}_1 = \bar{r}_2^T \cdot \bar{r}_2 \quad (12)$$

where  $\bar{a}_1$  and  $\bar{a}_2$  constitutes the optimal affine solution, therefore

$$N_{tr} = \min_{\bar{r}_1, \bar{r}_2 \in \mathcal{R}^3} \|P^+ \bar{x} - \bar{r}_1\|^2 + \|P^+ \bar{y} - \bar{r}_2\|^2 \quad \text{s.t.} \quad \bar{r}_1^T \cdot \bar{r}_2 = 0, \quad \bar{r}_1^T \cdot \bar{r}_1 = \bar{r}_2^T \cdot \bar{r}_2 \quad (13)$$

In Section 5 we present a closed-form solution for this metric, and in Section 6 we show how this metric can be used to bound the *image metric* from both above and below.

## 4 Summary of results

In the rest of the paper we prove the following results:

### 4.1 Transformation space:

The *transformation metric* defined in Eq. (13) has the following solution

$$N_{tr} = \frac{1}{2} \left( \bar{x}^T B \bar{x} + \bar{y}^T B \bar{y} - 2\sqrt{\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2} \right)$$

where  $B$  is defined in Eq. (10), and  $\bar{x}, \bar{y}$  in Eq. (5). The **best view** according to this metric is given by

$$\begin{aligned} \bar{x} &= PP^+(\beta_1 \bar{x} + \beta_2 \bar{y}) \\ \bar{y} &= PP^+(\gamma_1 \bar{x} + \gamma_2 \bar{y}) \end{aligned}$$

where  $\beta_1, \beta_2, \gamma_1, \gamma_2$  are defined in Appendix D.

## 4.2 Image space:

Using  $N_{tr}$  we can bound the *image metric* from both above and below. Denote

$$N_{af} = \|(I - PP^+)\tilde{x}\|^2 + \|(I - PP^+)\tilde{y}\|^2$$

we show that

$$N_{af} + \lambda_1 N_{tr} \leq N_{im} \leq N_{af} + \lambda_3 N_{tr}$$

where  $\lambda_1 \leq \lambda_2 \leq \lambda_3$  are the eigenvalues of  $P^T P$ . A sub-optimal solution to  $N_{im}$  is given by

$$N_{af} + \frac{2\mu_1\mu_2}{\mu_1 + \mu_2} N_{tr}$$

where the computation of  $\mu_1, \mu_2$  is described in Appendix E. A tighter upper bound is deduced from this sub-optimal solution

$$N_{im} \leq N_{af} + \text{H.M.}\{\lambda_2, \lambda_3\} N_{tr} \leq N_{af} + 2\lambda_2 N_{tr}$$

where  $\text{H.M.}\{\lambda_2, \lambda_3\} = \frac{2}{\frac{1}{\lambda_2} + \frac{1}{\lambda_3}}$  is the Harmonic mean of  $\lambda_2, \lambda_3$ . The sub-optimal solution is proposed as an initial guess for an iterative algorithm to compute  $N_{im}$ .

## 5 Closed-form solution in transformation space

We now present a metric to compare between 3D models and 2D images under weak perspective projection. The metric is a closed-form solution to the *transformation metric*,  $N_{tr}$  defined in Eq. (13). We use the notation developed in Section 3.  $B$  is the  $n \times n$  **characteristic matrix** of the object,  $\tilde{x}, \tilde{y} \in \mathcal{R}^n$  contain the  $x$ - and  $y$ -coordinates of the image features. The metric is given by

$$N_{tr} = \frac{1}{2} \left( \tilde{x}^T B \tilde{x} + \tilde{y}^T B \tilde{y} - 2\sqrt{\tilde{x}^T B \tilde{x} \cdot \tilde{y}^T B \tilde{y} - (\tilde{x}^T B \tilde{y})^2} \right) \quad (14)$$

This metric penalizes for the nonrigidities of the optimal affine transformation. Note that  $N_{tr} = 0$  if the two rigid constraints in Eq. (11) are satisfied. Otherwise,  $N_{tr} > 0$  represents the optimal penalty for a deviation from satisfying the two constraints.

### Derivation of the results:

In the rest of this section we prove that the expression for  $N_{tr}$ , given by Eq. (14), is indeed the solution to the *transformation metric* defined in Eq. (13). The proof proceeds as follows: Theorem 1 computes the minimal solution when  $\tilde{r}_1$  and  $\tilde{r}_2$  are restricted to the plane spanned by  $\tilde{a}_1$  and  $\tilde{a}_2$ ; Theorem 2 extends this result to three-space.

**Theorem 1:** When  $\tilde{r}_1$  and  $\tilde{r}_2$  are limited to  $\text{span}\{\tilde{a}_1, \tilde{a}_2\}$ ,  $N_{tr}$  is given by Eq. (14).

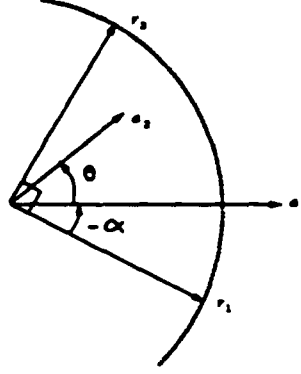


Figure 2: The vectors  $\bar{a}_1$ ,  $\bar{a}_2$ ,  $\bar{r}_1$ , and  $\bar{r}_2$  in the coordinate system specified in Theorem 1.  $\bar{a}_1$  and  $\bar{a}_2$  represent the solution for the affine case.  $\bar{r}_1$  and  $\bar{r}_2$  are constrained to be in the same plane with  $\bar{a}_1$  and  $\bar{a}_2$ , to be orthogonal, and to share the same norm.

**Proof:** We first define a new coordinate system in which

$$\begin{aligned}\bar{a}_1 &= w_1(1, 0) \\ \bar{a}_2 &= w_2(\cos \theta, \sin \theta) \\ \bar{r}_1 &= s(\cos \alpha, -\sin \alpha) \\ \bar{r}_2 &= s(\sin \alpha, \cos \alpha)\end{aligned}$$

(see Figure 2).  $\theta$  is the angle between  $\bar{a}_1$  and  $\bar{a}_2$ ,  $w_1$  and  $w_2$  are the lengths of  $\bar{a}_1$  and  $\bar{a}_2$  respectively.  $s$  is the common length of the two rotation vectors,  $\bar{r}_1$  and  $\bar{r}_2$ , and  $-\alpha$  is the angle between  $\bar{a}_1$  and  $\bar{r}_1$ . Without loss of generality it is assumed below that  $0^\circ \leq \theta \leq 180^\circ$  and  $-90^\circ \leq \alpha \leq 90^\circ$ . Notice that  $w_1$ ,  $w_2$ , and  $\theta$  are given and that  $s$  and  $\alpha$  are unknown.

Denote  $f$  the term to be minimized, that is

$$f(\alpha, s) = \|\bar{a}_1 - \bar{r}_1\|^2 + \|\bar{a}_2 - \bar{r}_2\|^2$$

then

$$\begin{aligned}f(\alpha, s) &= (w_1 - s \cos \alpha)^2 + s^2 \sin^2 \alpha + (s \sin \alpha - w_2 \cos \theta)^2 + (s \cos \alpha - w_2 \sin \theta)^2 \\ &= w_1^2 + w_2^2 + 2s^2 - 2s[(w_1 + w_2 \sin \theta) \cos \alpha + w_2 \cos \theta \sin \alpha]\end{aligned}$$

The partial derivatives of  $f$  are given by

$$\begin{aligned} f_{\alpha} &= 2s([w_1 + w_2 \sin \theta] \sin \alpha - w_2 \cos \theta \cos \alpha) \\ f_s &= 4s - 2([w_1 + w_2 \sin \theta] \cos \alpha + w_2 \cos \theta \sin \alpha) \end{aligned}$$

To find possible minima we equate these derivatives to zero

$$\begin{aligned} f_{\alpha} &= 0 \\ f_s &= 0 \end{aligned}$$

Solutions with  $s = 0$  are not optimal. In this case  $f(\alpha, 0) = w_1^2 + w_2^2$ , and later we show that solutions with  $s > 0$  always imply smaller values for  $f$ .

When  $s \neq 0$ ,  $f_{\alpha} = 0$  implies

$$\tan \alpha^{\min} = \frac{w_2 \cos \theta}{w_1 + w_2 \sin \theta}$$

therefore

$$\cos \alpha^{\min} = \frac{1}{\sqrt{1 + (\tan \alpha^{\min})^2}} = \frac{w_1 + w_2 \sin \theta}{\sqrt{w_1^2 + w_2^2 + 2w_1 w_2 \sin \theta}}$$

$f_s = 0$  implies

$$s^{\min} = \frac{1}{2}([w_1 + w_2 \sin \theta] \cos \alpha^{\min} + w_2 \cos \theta \sin \alpha^{\min})$$

Notice the similarity of this expression to the expression for  $f$ . At the minimum point  $f$  can be rewritten as

$$f^{\min} = w_1^2 + w_2^2 - 2(s^{\min})^2 \quad (15)$$

(From which it is apparent that any solution for  $f$  with  $s \neq 0$  would be smaller than the solution with  $s = 0$ .) Substituting for  $\alpha^{\min}$  we obtain

$$\begin{aligned} s^{\min} &= \frac{1}{2}([w_1 + w_2 \sin \theta] \cos \alpha^{\min} + w_2 \cos \theta \sin \alpha^{\min}) \\ &= \frac{1}{2} \cos \alpha^{\min} (w_1 + w_2 \sin \theta + w_2 \cos \theta \tan \alpha^{\min}) \\ &= \frac{w_1 + w_2 \sin \theta}{2\sqrt{w_1^2 + w_2^2 + 2w_1 w_2 \sin \theta}} \left( w_1 + w_2 \sin \theta + \frac{w_2^2 \cos^2 \theta}{w_1 + w_2 \sin \theta} \right) \\ &= \frac{1}{2} \sqrt{w_1^2 + w_2^2 + 2w_1 w_2 \sin \theta} \end{aligned}$$

and therefore

$$f^{\min} = w_1^2 + w_2^2 - 2(s^{\min})^2 = w_1^2 + w_2^2 - \frac{1}{2} (w_1^2 + w_2^2 + 2w_1 w_2 \sin \theta)$$

or,

$$f^{min} = \frac{1}{2}(w_1^2 + w_2^2 - 2w_1w_2 \sin \theta)$$

Recall that  $w_1$  and  $w_2$  are the lengths of  $\vec{a}_1$  and  $\vec{a}_2$ , that is

$$\begin{aligned} w_1^2 &= \vec{a}_1^T \cdot \vec{a}_1 = \vec{x}^T B \vec{x} \\ w_2^2 &= \vec{a}_2^T \cdot \vec{a}_2 = \vec{y}^T B \vec{y} \end{aligned}$$

and  $\theta$  is the angle between the two vectors, namely

$$w_1 w_2 \sin \theta = \sqrt{w_1^2 w_2^2 (1 - \cos^2 \theta)} = \sqrt{\vec{x}^T B \vec{x} \cdot \vec{y}^T B \vec{y} - (\vec{x}^T B \vec{y})^2}$$

We obtain that

$$f^{min} = \frac{1}{2} \left( \vec{x}^T B \vec{x} + \vec{y}^T B \vec{y} - 2\sqrt{\vec{x}^T B \vec{x} \cdot \vec{y}^T B \vec{y} - (\vec{x}^T B \vec{y})^2} \right)$$

□

In Theorem 1 we proved that if  $\vec{r}_1$  and  $\vec{r}_2$  are restricted to the plane spanned by  $\vec{a}_1$  and  $\vec{a}_2$ , the metric  $N_{tr}$  is given by Eq. (14). In Theorem 2 below we prove that any other solution for  $\vec{r}_1$  and  $\vec{r}_2$  results in a larger value for  $f$ , and therefore the minimum for  $f$  is obtained inside the plane, implying that  $N_{tr}$  indeed is given by Eq. (14).

**Theorem 2:** *The optimal  $\vec{r}_1$  and  $\vec{r}_2$  lie in the plane spanned by  $\vec{a}_1$  and  $\vec{a}_2$ .*

**Proof:** Assume, by way of contradiction, that  $\vec{r}_1, \vec{r}_2 \notin \text{span}\{\vec{a}_1, \vec{a}_2\}$ ; we show that the corresponding value for  $f$  is not minimal.

Consider first the plane spanned by  $\vec{r}_2$  and  $\vec{a}_1$ , and assume, by way of contradiction, that  $\vec{r}_1 \notin \text{span}\{\vec{r}_2, \vec{a}_1\}$ ; we show that there exists a vector  $\vec{r}'_1$  such that

$$\begin{aligned} \|\vec{r}'_1\| &= \|\vec{r}_2\| \\ \vec{r}'_1 &\perp \vec{r}_2 \end{aligned}$$

and

$$\|\vec{r}'_1 - \vec{a}_1\| < \|\vec{r}_1 - \vec{a}_1\|$$

contradicting the optimality of  $f$ .

Assume  $\|\vec{r}_2\| = s$ , and denote by  $\vec{r}'_1$  a vector with length  $s$  in the direction  $(\vec{r}_2 \times \vec{a}_1) \times \vec{r}_2$ . This vector lies in  $\text{span}\{\vec{r}_2, \vec{a}_1\}$  and satisfies

$$\begin{aligned} \|\vec{r}'_1\| &= \|\vec{r}_2\| \\ \vec{r}'_1 &\perp \vec{r}_2 \end{aligned}$$

(There exist two such vectors, opposing in their direction. We consider the one nearest to  $\bar{a}_1$ .)  
We now show that

$$\|\bar{r}'_1 - \bar{a}_1\| < \|\bar{r}_1 - \bar{a}_1\|$$

Denote the angle between  $\bar{a}_1$  and  $\bar{r}'_1$  by  $\alpha$ , and denote the angle between  $\bar{r}'_1$  and  $\bar{r}_1$  by  $\beta$ . Also, denote  $w_1 = \|\bar{a}_1\|$  and  $s = \|\bar{r}_1\| = \|\bar{r}_2\| = \|\bar{r}'_1\|$ . We can rotate the coordinate system so as to obtain

$$\begin{aligned}\bar{r}'_1 &= s(1, 0, 0) \\ \bar{r}_2 &= s(0, 1, 0) \\ \bar{a}_1 &= w_1(\cos \alpha, \sin \alpha, 0) \\ \bar{r}_1 &= s(\cos \beta, 0, \sin \beta)\end{aligned}$$

Now,

$$\begin{aligned}\|\bar{r}'_1 - \bar{a}_1\|^2 &= (s - w_1 \cos \alpha)^2 + w_1^2 \sin^2 \alpha = w_1^2 + s^2 - 2sw_1 \cos \alpha \\ \|\bar{r}_1 - \bar{a}_1\|^2 &= (s \cos \beta - w_1 \cos \alpha)^2 + w_1^2 \sin^2 \alpha + s^2 \sin^2 \beta = w_1^2 + s^2 - 2sw \cos \alpha \cos \beta\end{aligned}$$

and therefore, when  $\alpha \neq 0^\circ$  and  $\beta \neq 0^\circ$  (when  $\beta = 0^\circ$ ,  $\bar{r}_1$  and  $\bar{r}'_1$  coincide.)

$$\|\bar{r}'_1 - \bar{a}_1\| < \|\bar{r}_1 - \bar{a}_1\|$$

contradicting the minimality property. Therefore,  $\bar{r}_1 \in \text{span}\{\bar{r}_2, \bar{a}_1\}$ . Similarly, it can be shown that  $\bar{r}_2 \in \text{span}\{\bar{r}_1, \bar{a}_2\}$ , therefore all four vectors  $\bar{a}_1$ ,  $\bar{a}_2$ ,  $\bar{r}_1$ , and  $\bar{r}_2$  lie in a single plane.

□

**Corollary 3:** The transformation metric is given by

$$N_{tr} = \frac{1}{2} \left( \bar{x}^T B \bar{x} + \bar{y}^T B \bar{y} - 2\sqrt{\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2} \right)$$

and the best view for this metric is

$$\begin{aligned}\bar{x}^* &= PP^+(\beta_1 \bar{x} + \beta_2 \bar{y}) \\ \bar{y}^* &= PP^+(\gamma_1 \bar{x} + \gamma_2 \bar{y})\end{aligned}$$

where

$$\begin{aligned}\beta_1 &= \frac{1}{2} \left( 1 + \frac{\bar{y}^T B \bar{y}}{\sqrt{\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2}} \right) \\ \beta_2 = \gamma_1 &= -\frac{\bar{x}^T B \bar{y}}{2\sqrt{\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2}} \\ \gamma_2 &= \frac{1}{2} \left( 1 + \frac{\bar{x}^T B \bar{x}}{\sqrt{\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2}} \right)\end{aligned}$$



**Proof:** The expression for the metric immediately follows from Theorem 1 and 2. The expression for the **best view** is developed in the Appendix D.

□

## 6 Solution in image space

In order to compute the *image metric* as defined in section 3, we need to solve the constraint minimization problem defined in Eq. (6)

$$N_{im} = \min_{\vec{r}_1, \vec{r}_2 \in \mathbb{R}^3} \|\vec{x} - P\vec{r}_1\|^2 + \|\vec{y} - P\vec{r}_2\|^2 \quad \text{s.t.} \quad \vec{r}_1^T \cdot \vec{r}_2 = 0, \quad \vec{r}_1^T \cdot \vec{r}_1 = \vec{r}_2^T \cdot \vec{r}_2$$

Section 6.1 shows that  $N_{tr}$ , computed in the previous section, can be used to bound  $N_{im}$  from both above and below. Section 6.2 describes a direct method to compute a sub-optimal approximation to  $N_{im}$  and outlines an iterative algorithm to improve this estimate to obtain the optimal  $N_{im}$ .

### 6.1 Bounding the image metric with the transformation metric

In this section we show that using the *transformation metric* defined in Section 5  $N_{tr}$ , and the affine metric  $N_{af}$  (given in Eq. (9)), we can bound the *image metric*  $N_{im}$  from both above and below. We prove the following theorem:

**Theorem 4:** Let  $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3$  denote the three eigenvalues of  $P^T P$ , then

$$N_{af} + \lambda_1 N_{tr} \leq N_{im} \leq N_{af} + \lambda_3 N_{tr} \quad (16)$$

**Proof:** Denote by  $\vec{r}_1^*$  and  $\vec{r}_2^*$  the vectors that minimize the term for the *image metric* given in Eq. (6), namely

$$N_{im} = \|\vec{x} - P\vec{r}_1^*\|^2 + \|\vec{y} - P\vec{r}_2^*\|^2$$

and denote by  $\vec{r}_1$  and  $\vec{r}_2$  the vectors that minimize the *transformation metric* given in Eq. (13), namely

$$N_{tr} = \|P^+ \vec{x} - \vec{r}_1\|^2 + \|P^+ \vec{y} - \vec{r}_2\|^2$$

We start by showing the upper bound. Since  $\vec{r}_1^*$  and  $\vec{r}_2^*$  minimize the term for  $N_{im}$ , we can write

$$\begin{aligned} N_{im} &= \|\vec{x} - P\vec{r}_1^*\|^2 + \|\vec{y} - P\vec{r}_2^*\|^2 \\ &\leq \|\vec{x} - P\vec{r}_1\|^2 + \|\vec{y} - P\vec{r}_2\|^2 \end{aligned}$$

We now break each term in this sum into two orthogonal components as follows

$$\vec{x} - P\vec{r}_1 = (\vec{x} - PP^+ \vec{x}) + (PP^+ \vec{x} - P\vec{r}_1)$$

for which it holds that

$$(\bar{x} - PP^+\bar{x})^T \cdot (PP^+\bar{x} - P\bar{r}_1) = 0$$

The orthogonality readily follows from the identity

$$(PP^+)^T P = (P^+)^T P^T P = P(P^T P)^{-1}(P^T P) = P$$

Since the two components are orthogonal it holds that

$$\|\bar{x} - P\bar{r}_1\|^2 = \|\bar{x} - PP^+\bar{x}\|^2 + \|PP^+\bar{x} - P\bar{r}_1\|^2$$

and, similarly,

$$\|\bar{y} - P\bar{r}_2\|^2 = \|\bar{y} - PP^+\bar{y}\|^2 + \|PP^+\bar{y} - P\bar{r}_2\|^2$$

Therefore (recall that  $\bar{r}_1$  and  $\bar{r}_2$  minimize  $N_{tr}$  and that  $\lambda_3$  is the largest eigenvalue of  $P^T P$ )

$$\begin{aligned} N_{im} &\leq \|\bar{x} - P\bar{r}_1\|^2 + \|\bar{y} - P\bar{r}_2\|^2 \\ &= \|\bar{x} - PP^+\bar{x}\|^2 + \|PP^+\bar{x} - P\bar{r}_1\|^2 + \|\bar{y} - PP^+\bar{y}\|^2 + \|PP^+\bar{y} - P\bar{r}_2\|^2 \\ &= \|(I - PP^+)\bar{x}\|^2 + \|(I - PP^+)\bar{y}\|^2 + \|P(P^+\bar{x} - \bar{r}_1)\|^2 + \|P(P^+\bar{y} - \bar{r}_2)\|^2 \\ &= N_{af} + \|P(P^+\bar{x} - \bar{r}_1)\|^2 + \|P(P^+\bar{y} - \bar{r}_2)\|^2 \\ &\leq N_{af} + \lambda_3(\|(P^+\bar{x} - \bar{r}_1)\|^2 + \|(P^+\bar{y} - \bar{r}_2)\|^2) \\ &= N_{af} + \lambda_3 N_{tr} \end{aligned}$$

Next, we prove the lower bound. The proof is similar to the proof in the upper bound case, but this time we start by breaking up the terms into orthogonal components. Then we use the facts that  $\bar{r}_1$  and  $\bar{r}_2$  minimize  $N_{tr}$  and that  $\lambda_1$  is the smallest eigenvalue of  $P^T P$ .

$$\begin{aligned} N_{im} &= \|\bar{x} - P\bar{r}_1\|^2 + \|\bar{y} - P\bar{r}_2\|^2 \\ &= \|\bar{x} - PP^+\bar{x}\|^2 + \|PP^+\bar{x} - P\bar{r}_1\|^2 + \|\bar{y} - PP^+\bar{y}\|^2 + \|PP^+\bar{y} - P\bar{r}_2\|^2 \\ &= \|(I - PP^+)\bar{x}\|^2 + \|(I - PP^+)\bar{y}\|^2 + \|P(P^+\bar{x} - \bar{r}_1)\|^2 + \|P(P^+\bar{y} - \bar{r}_2)\|^2 \\ &= N_{af} + \|P(P^+\bar{x} - \bar{r}_1)\|^2 + \|P(P^+\bar{y} - \bar{r}_2)\|^2 \\ &\geq N_{af} + \lambda_1(\|(P^+\bar{x} - \bar{r}_1)\|^2 + \|(P^+\bar{y} - \bar{r}_2)\|^2) \\ &\geq N_{af} + \lambda_1 N_{tr} \end{aligned}$$

Consequently

$$N_{af} + \lambda_1 N_{tr} \leq N_{im} \leq N_{af} + \lambda_3 N_{tr}$$

□

## 6.2 Direct solution for the image metric

In this section we develop tighter bounds on the *image metric* by direct methods, following the same steps we took in the derivation of the *transformation metric* in Section 5. Unlike for the *transformation metric*, we cannot obtain a closed-form solution for the *image metric*, but we can obtain a better estimator than we have previously described. This also enables us to develop an iterative method to compute the distance exactly.

In section 6.2.1 we describe a change of coordinate system, arriving at a minimization problem which is similar to the one we had to solve for the *transformation metric*. The difference is that the sought vectors are constrained to lie on an ellipsoid rather than a sphere, and the ellipsoid is defined by a  $3 \times 3$  positive-definite version of the characteristic matrix  $B$ .

In section 6.2.2 we restrict the solution vectors,  $\tilde{u}, \tilde{v}$ , to lie in a plane with the data vectors,  $\tilde{x}, \tilde{y}$  and we derive the optimal solution under this constraint. The solution, however, is only sub-optimal, since in contrast to the *transformation metric*, the optimal solution in this case does not have to lie in the plane. Using this solution we derive a tighter upper bound on the optimal solution.

In section 6.2.3 we describe the general problem that needs to be solved, and outline an iterative method. We propose the solution obtained in the plane as an initial guess for this method.

### 6.2.1 Reducing the dimensionality of the problem

In Section 6.1 we have shown that the *image metric* can be broken into two orthogonal terms, implying that

$$N_{im} = N_{af} + \|P(P^+ \tilde{x} - \tilde{r}_1)\|^2 + \|P(P^+ \tilde{y} - \tilde{r}_2)\|^2 \quad (17)$$

This property is useful for a direct computation of the *image metric*. The first term,  $N_{af}$ , does not depend on  $\tilde{r}_1, \tilde{r}_2$ . To compute  $N_{im}$ , therefore, only the second term needs to be minimized

$$\min_{\tilde{r}_1, \tilde{r}_2 \in \mathbb{R}^3} \|PP^+ \tilde{x} - P\tilde{r}_1\|^2 + \|PP^+ \tilde{y} - P\tilde{r}_2\|^2 \quad \text{s.t.} \quad \tilde{r}_1^T \cdot \tilde{r}_2 = 0, \quad \tilde{r}_1^T \cdot \tilde{r}_1 = \tilde{r}_2^T \cdot \tilde{r}_2 \quad (18)$$

Note first that  $PP^+ \tilde{x}$  and  $PP^+ \tilde{y}$ , two vectors in  $\mathbb{R}^n$ , both lie in a single linear subspace of dimension 3. (This follows from the fact, shown in [Ullman and Basri, 1991], that every image of a 3D object can be written as a linear combination of three independent views.) Moreover, the three columns of  $P$  lie in the same subspace. It therefore follows that the vectors  $\tilde{u} = P\tilde{r}_1$  and  $\tilde{v} = P\tilde{r}_2$  must also lie in this subspace.

Denote  $\tilde{X} = PP^+ \tilde{x}$  and  $\tilde{Y} = PP^+ \tilde{y}$ , the projection of  $\tilde{x}$  and  $\tilde{y}$  to the column space of  $P$ , and denote  $\tilde{u} = P\tilde{r}_1$  and  $\tilde{v} = P\tilde{r}_2$ . (Note that  $\tilde{r}_1 = P^+ \tilde{u}$ ,  $\tilde{r}_2 = P^+ \tilde{v}$ , and  $B = (P^+)^T P^+$ , the characteristic matrix of the object.) We rewrite the problem as follows

$$\min_{\tilde{u}, \tilde{v} \in \mathbb{R}^n} \|\tilde{X} - \tilde{u}\|^2 + \|\tilde{Y} - \tilde{v}\|^2 \quad \text{s.t.} \quad \tilde{u}^T B \tilde{v} = 0, \quad \tilde{u}^T B \tilde{u} = \tilde{v}^T B \tilde{v} \quad (19)$$

Since all the vectors,  $\vec{X}$ ,  $\vec{Y}$ ,  $\vec{u}$ , and  $\vec{v}$ , lie in a 3D subspace (the column space of  $P$ ) we can perform the minimization in  $\mathcal{R}^3$ . To transform the system into  $\mathcal{R}^3$ , we rotate the vectors and the characteristic matrix  $B$  so as to get nontrivial (nonzero) values only in three of the coordinates. Recall that distances and quadratic forms are invariant under rotation. The rotation matrix  $\Omega$  that should be applied to all terms is defined by the eigenvectors of  $B$ . Applying this matrix to  $B$  (in the form  $\Omega^T B \Omega$ ) results in a diagonal matrix with the three positive eigenvalues of  $B$ .

### 6.2.2 Closed-form solution in the plane

**Theorem 5:** When  $\vec{u}$  and  $\vec{v}$  are limited to  $\text{span}\{\vec{X}, \vec{Y}\}$ , the solution of Eq. (19) is given by

$$\tilde{N}_{im} = \frac{\mu_1 \mu_2}{\mu_1 + \mu_2} \left( \vec{x}^T B \vec{x} + \vec{y}^T B \vec{y} - 2\sqrt{\vec{x}^T B \vec{x} \cdot \vec{y}^T B \vec{y} - (\vec{x}^T B \vec{y})^2} \right) \quad (20)$$

where  $\sqrt{\mu_1} \leq \sqrt{\mu_2}$  are the principal axes of the ellipse, defined by the intersection of the ellipsoid  $B$  with the plane  $\text{span}\{\vec{X}, \vec{Y}\}$ .

Note the similarity between this solution and  $N_{tr}$  in Eq. (14). In fact,

$$\tilde{N}_{im} = \frac{2\mu_1 \mu_2}{\mu_1 + \mu_2} N_{tr} \quad (21)$$

The proof closely follows the proof for  $N_{tr}$  presented in Section 5 (Theorem 2). We therefore skip some of the details.

**Proof:** We first define a new coordinate system in which

$$\begin{aligned} \vec{X} &= w_1(\sqrt{\mu_1} \cos \eta, \sqrt{\mu_2} \sin \eta) \\ \vec{Y} &= w_2(\sqrt{\mu_1} \cos \theta, \sqrt{\mu_2} \sin \theta) \\ \vec{u} &= s(\sqrt{\mu_1} \cos \alpha, \sqrt{\mu_2} \sin \alpha) \\ \vec{v} &= s(-\sqrt{\mu_1} \sin \alpha, \sqrt{\mu_2} \cos \alpha) \\ B &= \begin{pmatrix} \frac{1}{\mu_1} & 0 & B_{13} \\ 0 & \frac{1}{\mu_2} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{pmatrix} \end{aligned}$$

Without loss of generality it is assumed below that  $-90^\circ \leq \eta \leq 90^\circ$ ,  $\eta \leq \theta \leq \eta + 180^\circ$ , and  $-90^\circ \leq \alpha \leq 90^\circ$ . Notice that  $w_1$ ,  $w_2$ ,  $\eta$  and  $\theta$  are given and that  $s$  and  $\alpha$  are unknown.

Notice that this setting of coordinate system is similar to the one used in Theorem 1 with the exceptions that here  $\vec{u}$  and  $\vec{v}$  lie on an ellipse rather than on a circle, and that in general none of the points can be brought to lie on a principal axis.

Denote by  $f$  the term to be minimized, that is

$$f(\alpha, s) = \|\vec{X} - \vec{u}\|^2 + \|\vec{Y} - \vec{v}\|^2$$

then

$$\begin{aligned}
f(\alpha, s) &= \mu_1(w_1 \cos \eta - s \cos \alpha)^2 + \mu_2(w_1 \sin \eta - s \sin \alpha)^2 + \mu_1(w_2 \cos \theta + s \sin \alpha)^2 + \\
&\quad \mu_2(w_2 \sin \theta - s \cos \alpha)^2 \\
&= w_1^2(\mu_1 \cos^2 \eta + \mu_2 \sin^2 \eta) + w_2^2(\mu_1 \cos^2 \theta + \mu_2 \sin^2 \theta) + s^2(\mu_1 + \mu_2) - \\
&\quad 2s(w_1\mu_1 \cos \eta \cos \alpha + w_1\mu_2 \sin \eta \sin \alpha - w_2\mu_1 \cos \theta \sin \alpha + w_2\mu_2 \sin \theta \cos \alpha)
\end{aligned}$$

The partial derivatives of  $f$  are given by

$$\begin{aligned}
f_\alpha &= 2s[(w_1\mu_1 \cos \eta + w_2\mu_2 \sin \theta) \sin \alpha - (w_1\mu_2 \sin \eta - w_2\mu_1 \cos \theta) \cos \alpha] \\
f_s &= 2s(\mu_1 + \mu_2) - 2(w_1\mu_1 \cos \eta \cos \alpha + w_1\mu_2 \sin \eta \sin \alpha - w_2\mu_1 \cos \theta \sin \alpha + w_2\mu_2 \sin \theta \cos \alpha)
\end{aligned}$$

To find possible minima we equate these derivatives to zero

$$\begin{aligned}
f_\alpha &= 0 \\
f_s &= 0
\end{aligned}$$

Again, solutions with  $s = 0$  can be ignored since they do not correspond to the global minimum (for a similar reason as in the proof of Theorem 1).

When  $s \neq 0$ ,  $f_\alpha = 0$  implies

$$\tan \alpha^{\min} = \frac{w_1\mu_2 \sin \eta - w_2\mu_1 \cos \theta}{w_1\mu_1 \cos \eta + w_2\mu_2 \sin \theta}$$

$f_s = 0$  implies

$$s^{\min} = \frac{w_1\mu_1 \cos \eta + w_2\mu_2 \sin \theta}{\cos \alpha^{\min}(\mu_1 + \mu_2)}$$

and, similarly to Eq. (15),

$$f^{\min} = w_1^2(\mu_1 \cos^2 \eta + \mu_2 \sin^2 \eta) + w_2^2(\mu_1 \cos^2 \theta + \mu_2 \sin^2 \theta) - (\mu_1 + \mu_2)(s^{\min})^2 \quad (22)$$

We substitute  $s^{\min}$  and  $\cos \alpha^{\min}$ , using the identity  $\cos \alpha = \frac{1}{\sqrt{1+\tan^2 \alpha}}$ , into Eq. (22). After some manipulations, we obtain

$$\tilde{N}_{im} = f^{\min} = \frac{\mu_1\mu_2}{\mu_1 + \mu_2} (w_1^2 + w_2^2 - 2w_1w_2 \sin(\theta - \eta)) \quad (23)$$

Note that

$$(PP^+)^T B (PP^+) = (P^+)^T P^T (P^+)^T P^+ PP^+ = (P^+)^T (P^+ P)^T (P^+ P) P^+ = (P^+)^T P^+ = B \quad (24)$$

from which it follows that

$$\begin{aligned} w_1^2 &= \bar{X}^T B \bar{X} = \bar{x}^T B \bar{x} \\ w_2^2 &= \bar{Y}^T B \bar{Y} = \bar{y}^T B \bar{y} \\ w_1 w_2 \cos(\theta - \eta) &= \bar{X}^T B \bar{Y} = \bar{x}^T B \bar{y} \end{aligned} \quad (25)$$

We substitute the identities from Eq. (25) into Eq. (23), obtaining the expression for  $\tilde{N}_{im}$  in Eq. (20).  
□

The derivation for  $\mu_1$  and  $\mu_2$  is given in Appendix E.

The sub-optimal solution in the plane can be used to improve the bounds on the *image metric*, which were previously discussed in Theorem 4.

**Theorem 6:** Let  $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3$  be the three eigenvalues of  $P^T P$ , then

$$N_{af} + \lambda_1 N_{tr} \leq N_{im} \leq N_{af} + \text{H.M.}\{\lambda_2, \lambda_3\} N_{tr} \quad (26)$$

where  $\text{H.M.}\{\lambda_2, \lambda_3\} = \frac{2}{\frac{1}{\lambda_2} + \frac{1}{\lambda_3}}$ , the Harmonic Mean of  $\lambda_2, \lambda_3$ .

**Proof:** The eigenvalues of the characteristic matrix  $B$  are  $\frac{1}{\lambda_1}$ ,  $\frac{1}{\lambda_2}$ , and  $\frac{1}{\lambda_3}$ . (This is shown in Appendix E.) Since  $1/\mu_1$  and  $1/\mu_2$  represent the eigenvalues of a section of  $B$  it holds that (see, e.g., [Strang, 1976] p. 270)

$$\frac{1}{\lambda_1} \geq \frac{1}{\mu_1} \geq \frac{1}{\lambda_2} \geq \frac{1}{\mu_2} \geq \frac{1}{\lambda_3}$$

Using Eq. 21 we obtain that

$$\tilde{N}_{im} = \frac{2\mu_1\mu_2}{\mu_1 + \mu_2} N_{tr} = \frac{2}{\frac{1}{\mu_1} + \frac{1}{\mu_2}} N_{tr} \leq \frac{2}{\frac{1}{\lambda_2} + \frac{1}{\lambda_3}} N_{tr} = \text{H.M.}\{\lambda_2, \lambda_3\} N_{tr}$$

And, using Eq. 17 we obtain the upper bound

$$N_{im} \leq N_{af} + \tilde{N}_{im} \leq N_{af} + \text{H.M.}\{\lambda_2, \lambda_3\} N_{tr}$$

□

**Corollary 7:**

$$N_{af} + \lambda_1 N_{tr} \leq N_{im} \leq N_{af} + \frac{2\mu_1\mu_2}{\mu_1 + \mu_2} N_{tr} \quad (27)$$

Note that, since  $H.M.\{a, b\} \leq 2 \min\{a, b\}$  for every  $a, b$ , we have the following corollary.

**Corollary 8:**

$$N_{af} + \lambda_1 N_{tr} \leq N_{im} \leq N_{af} + 2\lambda_2 N_{tr}$$

We cannot yet improve the lower bound in theorem 4; but we conjecture that

**Conjecture 1:** Let  $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3$  be the three eigenvalues of  $P^T P$ , then

$$N_{af} + H.M.\{\lambda_1, \lambda_2\} N_{tr} \leq N_{im} \leq N_{af} + H.M.\{\lambda_2, \lambda_3\} N_{tr} \quad (28)$$

**Motivation:** We know that if the two data points  $\vec{X}, \vec{Y}$  lie on the ellipse whose principal axes are of length  $\lambda_1, \lambda_2$  (the smallest cross-section of the ellipsoid  $B$ ), then

$$N_{im} = N_{af} + H.M.\{\lambda_1, \lambda_2\} N_{tr}$$

We can show that this solution is a local minimum, namely, it is not possible to improve the solution by applying small perturbations to the solution vectors.

□

### 6.2.3 An iterative optimal solution

The solution we obtained in Theorem 5 is sub-optimal; it is not the lowest distance. We now give the cost function, a function of four variables, which should be minimized to obtain the precise value of the *image metric*.

We first define a coordinate system such that

$$\begin{aligned} \vec{X} &= w_1(\sqrt{\lambda_1} \cos \theta \cos \nu, \sqrt{\lambda_2} \cos \theta \sin \nu, \sqrt{\lambda_3} \sin \theta) \\ \vec{Y} &= w_2(\sqrt{\lambda_1} \cos \zeta \cos \eta, \sqrt{\lambda_2} \cos \zeta \sin \eta, \sqrt{\lambda_3} \sin \zeta) \\ \vec{u} &= s(\sqrt{\lambda_1} \cos \alpha \cos \beta, \sqrt{\lambda_2} \cos \alpha \sin \beta, \sqrt{\lambda_3} \sin \alpha) \\ \vec{v} &= s(\sqrt{\lambda_1}(\sin \beta \cos \gamma + \sin \alpha \cos \beta \sin \gamma), \sqrt{\lambda_2}(-\cos \beta \cos \gamma + \sin \alpha \sin \beta \sin \gamma), \\ &\quad -\sqrt{\lambda_3} \cos \alpha \sin \gamma) \\ B &= \begin{pmatrix} \frac{1}{\lambda_1} & 0 & 0 \\ 0 & \frac{1}{\lambda_2} & 0 \\ 0 & 0 & \frac{1}{\lambda_3} \end{pmatrix} \end{aligned}$$

where  $w_1, w_2, \lambda_1, \lambda_2, \lambda_3, \zeta, \eta, \theta$ , and  $\nu$  are known, and  $s, \alpha, \beta$  and  $\gamma$  are free.

Note that this setting of coordinate system is similar to the one used in Theorem 5, but now  $\vec{u}$  and  $\vec{v}$  lie on an ellipsoid rather than on an ellipse.

In this notation the free parameters are selected so as to satisfy the two rigid constraints,  $\bar{u}^T B \bar{u} = \bar{v}^T B \bar{v}$  and  $\bar{u}^T B \bar{v} = 0$ . To compute the *image metric*, the following function should be minimized.

$$\begin{aligned}
f(s, \alpha, \beta, \gamma) = & \lambda_1(s \cos \alpha \cos \beta - w_1 \cos \theta \cos \nu)^2 + \lambda_2(s \cos \alpha \sin \beta - w_1 \cos \theta \sin \nu)^2 + \\
& \lambda_3(s \sin \alpha - w_1 \sin \theta)^2 + \\
& \lambda_1(s \sin \beta \cos \gamma + s \sin \alpha \cos \beta \sin \gamma - w_2 \cos \zeta \cos \eta)^2 + \\
& \lambda_2(-s \cos \beta \cos \gamma + s \sin \alpha \sin \beta \sin \gamma - w_2 \cos \zeta \sin \eta)^2 + \\
& \lambda_3(-s \cos \alpha \sin \gamma - w_2 \sin \zeta)^2
\end{aligned} \tag{29}$$

$N_{im}$  is the global minimum of  $f(s, \alpha, \beta, \gamma)$ . Assuming that  $f(s, \alpha, \beta, \gamma)$  is convex in the area that contains both the global minimum  $N_{im}$  and the sub-optimal solution  $(N_{af} + \tilde{N}_{im})$ , we can employ the following iterative method to compute  $N_{im}$ :

1. compute  $\tilde{N}_{im}$ ;
2. improve the solution by any gradient-descent method until a local minimum is obtained.

If the convexity assumption is correct, this method returns the correct *image metric*, otherwise it may return a sub-optimal solution.

## 7 Simulations

To test the presented metric we have compared it with the alignment method. As was mentioned in Section 2 the alignment method involves the selection of a small subset of correspondences (*alignment key*), solving for the transformation using this subset, and then transforming the rest of the points and measuring their distance from the corresponding image points. The obtained distance critically depends on the choice of alignment key. Different choices produce different distance measures between the model and the image. The results are almost always sub-optimal, since it is usually better to match all points with small errors than to exactly match a subset of points and project the errors entirely onto the others.

In our simulations, models composed of four points were projected to the image using weak perspective projection. Gaussian noise (with standard deviation 0.05 of the radius of the 3D object) was added to the obtained images. Using the expression for  $N_{tr}$  given in (14), we computed the upper and lower bounds on the *image metric* between the model to the noisy images. In addition, we computed the corresponding *alignment distances*, each reflecting the distance between one model point and its predicted projection in the image after the alignment of the remaining three image points to the model.

The figures below summarize our results. Figure 3 shows the percentage of *alignment distances* which actually lie within the bounds on the *image metric* computed by our metric



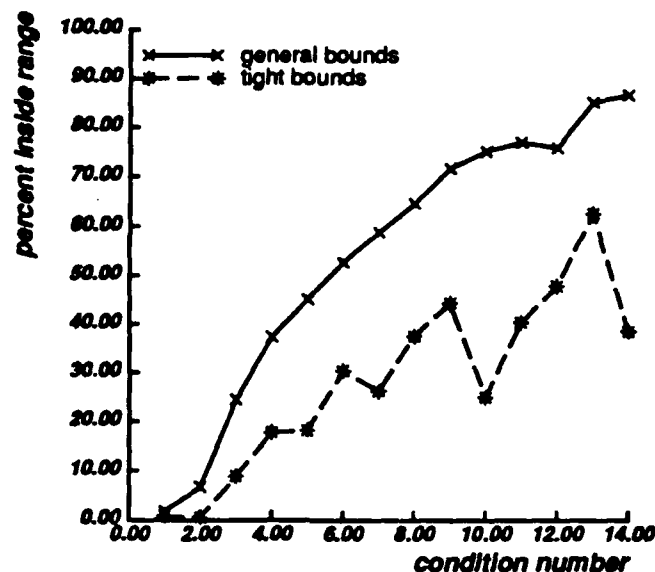


Figure 3: The percent of *alignment distances* which lie within the bounds on the *image metric* computed from our closed-form equations. The abscissa gives the condition number of the characteristic matrix,  $B$ , which determines how far apart the lower and upper bounds on the *image metric* are. The larger the condition number is, the further apart the bounds are. Solid graph: alignment distances relative to the wide bounds from Eq. (26). Dashed lines: alignment distances relative to the tight upper bound from Eq. (27).

(given in Eq. (26)). It can be seen that when the bounds are relatively tight (when the condition number on the characteristic matrix  $B$  is relatively low) most of the alignment solutions exceed the upper bound. Only when the condition number gets larger do the alignment distances lie within the bounds. When a tighter upper bound is used (Eq. (27)), a smaller portion of the alignment distances actually lie within the bounds.

Figure 4 shows the maximal and minimal *alignment distances* obtained in different runs relative to the upper and lower bounds on the *image metric*, given in Eq. (26) and Eq. (27). It can be seen that in many cases even the best alignment solution (the one that minimizes the distance) still exceeds the upper bound.

## 8 Summary

We have proposed a *transformation metric* to measure the similarity between 3D models and 2D images. The *transformation metric* measures the amount of affine deformation applied to the object to produce the given image. A simple, closed-form solution for this metric has been presented. This solution is optimal in transformation space, and it is used to bound the *image metric* from both above and below.

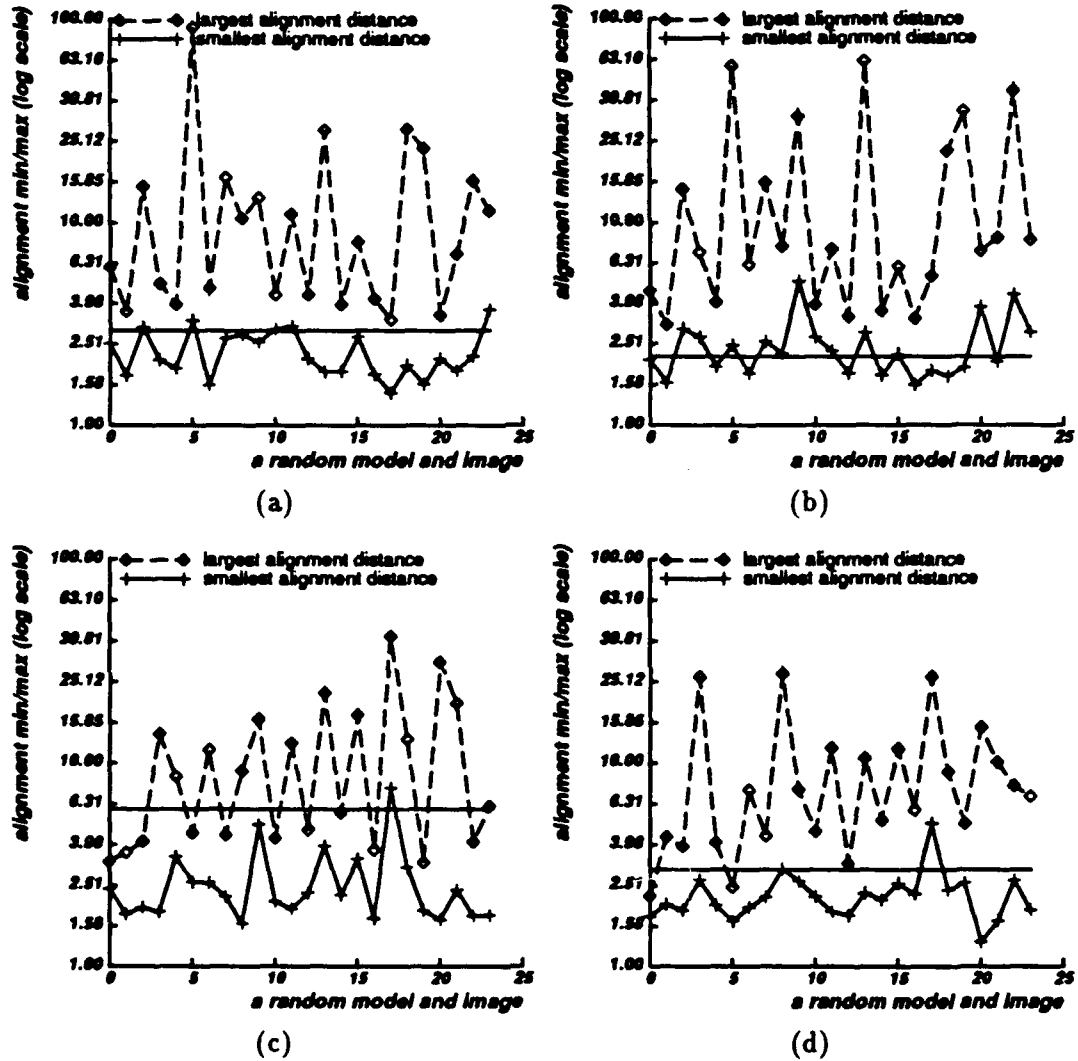


Figure 4: The maximal and minimal *alignment distances* are plotted for a number of models and objects, varying along the abscissa. The distances in these plots were normalized so as to obtain constant lower and upper bounds (the lower bound is set to 1; the upper bound is set to be the average ratio of the upper bound to the lower bound in each sequence of runs). Small (between 1.5 and 2.5) and large (between 4.5 and 5.5) condition numbers are used, and the results are compared to both the wide (Eq. (26)) and the tight (Eq. (27)) bounds. (a) Small condition number, wide bounds. (b) Small condition number, tight bounds. (c) Large condition number, wide bounds. (d) Large condition number, tight bounds.

The *transformation metric* presented in this paper can be used in several different ways in the recognition and classification tasks:

1. It provides a direct assessment of the similarity between models and images. Measuring the amount of deformation applied to the objects makes it suited for the task of object classification where the uncertainty in the structure of the observed objects is inherent.
2. The *transformation metric* can be used to bound the *image metric*, the distance between the image and the closest view of the object, from both above and below. As shown by our simulations, these bounds often provide better estimates than those provided by using alignment. Consequently, we believe that in many cases the bounds suffice to unequivocally determine the identity of the observed object.
3. The *transformation metric* provides a sub-optimal closed-form estimate for the *image metric*. A scheme which uses this measure will prefer "symmetric" objects, objects whose convex-hull is close to a sphere, over other objects which are significantly stretched or contracted along one spatial dimension. This solution can also be used as an initial guess in an iterative process that computes the optimal value of the image metric numerically.

## Appendices

### A Metric properties

The measures described in this paper compare entities of different dimensionalities: 3D objects and 2D images. We define a *metric* for comparing such entities as follows. Let  $P$  be a set of  $n$  model points, and let  $q$  be a set of  $n$  corresponding image points. A distance function,  $N(P, q)$ , defined using a difference function  $\partial(q, q')$  between two views (see Section 3), is called a *metric* if

1.  $N(P, q) \geq 0$  for every model  $P$  and image  $q$ .
2.  $N(P, q) = 0$  if, and only if,  $q$  is a rigid view of  $P$ .
3.  $\forall q', N(P, q) \leq N(P, q') + \partial(q - q')$

For the *image metric*,  $N_{im}$ ,  $\partial$  is simply the Euclidean distance between corresponding points in the compared images. It is straightforward to see that the conditions hold for this case. In the rest of this appendix we prove that these conditions also hold for the *transformation metric*,  $N_{tr}$ .

### Transformation metric

The *transformation metric*,  $N_{tr}$ , measures the amount of "affine deformation" applied to the object in the image. The metric conditions for  $N_{tr}$  are defined as follows.

1.  $N(P, q) \geq 0$  for every model  $P$  and image  $q$ .
2.  $N(P, q) = 0$  if, and only if, there exists a rigid view which coincides with  $PP^+q$ . (In other words, the best affine view of the object is a rigid view and there is no "affine deformation".)
3.  $\forall q', N(P, q) \leq N(P, q') + \|P^+(q - q')\|$

**Theorem 9:**  $N_{tr}$  is a metric.

**Proof:**

1.  $N_{tr} \geq 0$ .  $N_{tr}$  minimizes a non-negative distance function. It is therefore always non-negative.
2.  $N_{tr} = 0$  if, and only if, the best affine view is rigid. Denote  $\bar{x}$  and  $\bar{y}$  the  $x$  and  $y$  coordinates of the points in  $q$ , according to Eq. (14)

$$\begin{aligned}
 N_{tr} &= 0 \\
 \iff (\bar{x}^T B \bar{x} + \bar{y}^T B \bar{y})^2 &= 4(\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2) \\
 \iff (\bar{x}^T B \bar{x})^2 + 2(\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y}) + (\bar{y}^T B \bar{y})^2 &= 4(\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y}) - 4(\bar{x}^T B \bar{y})^2 \\
 \iff (\bar{x}^T B \bar{x} - \bar{y}^T B \bar{y})^2 &= -4(\bar{x}^T B \bar{y})^2
 \end{aligned}$$

This equation holds if, and only if, both sides are zero implying that

$$\begin{aligned}
 \bar{x}^T B \bar{x} &= \bar{y}^T B \bar{y} \\
 \bar{x}^T B \bar{y} &= 0
 \end{aligned}$$

The best affine view of the object is given by  $PP^+x, PP^+y$ . Following Eq. (11), the best affine view also satisfies the rigidity constraints above, and therefore it forms a rigid view.

3. The metric  $N_{tr}$  is defined in Eq. (13) as:

$$N_{tr}(P, q) = \min_{\bar{r}_1, \bar{r}_2 \in \mathcal{R}^3} \|P^+ \bar{x} - \bar{r}_1\|^2 + \|P^+ \bar{y} - \bar{r}_2\|^2 \quad \text{s.t.} \quad \bar{r}_1^T \cdot \bar{r}_2 = 0, \quad \bar{r}_1^T \cdot \bar{r}_1 = \bar{r}_2^T \cdot \bar{r}_2$$

Let  $\bar{w}_1$  and  $\bar{w}_2$  be the optimal vectors for  $q'$ , that is

$$N_{tr}(P, q') = \|P^+ \bar{x}' - \bar{w}_1\|^2 + \|P^+ \bar{y}' - \bar{w}_2\|^2$$

And we obtain

$$\begin{aligned}
& N_{tr}(P, q') + \|P^+(q - q')\| \\
&= \|P^+ \bar{x}' - \bar{w}_1\|^2 + \|P^+ \bar{y}' - \bar{w}_2\|^2 + \|P^+ \bar{x} - P^+ \bar{x}'\|^2 + \|P^+ \bar{y} - P^+ \bar{y}'\|^2 \\
&\geq \|P^+ \bar{x} - \bar{w}_1\|^2 + \|P^+ \bar{y} - \bar{w}_2\|^2 \\
&\geq \min_{\bar{r}_1, \bar{r}_2 \in \mathcal{R}^3} \|P^+ \bar{x} - \bar{r}_1\|^2 + \|P^+ \bar{y} - \bar{r}_2\|^2 = N_{tr}(P, q)
\end{aligned}$$

## B The computation of the characteristic matrix

In Eq (10) the characteristic matrix  $B$  was defined using the matrix of *Euclidean* model point coordinates  $P$ . We now give a more general (though equivalent) definition of  $B$  using a matrix of *affine* model point coordinates  $Q$ . Namely, the point coordinates in  $Q$  are given in a coordinate system whose axes are not necessarily orthonormal. This definition makes it possible to compute  $B$  directly from three or more images with a completely *linear* algorithm, which requires no more than pseudo-inverse.

We select an affine coordinate system whose independent axes are defined by three of the object points, to be called the basis points. Let  $P_{bas}$  denote the submatrix of  $P$  corresponding to the coordinates of the basis points, and let  $Q$  denote the affine coordinates of all the object points in this basis. It immediately follows that:

$$P = Q \cdot P_{bas}$$

Let  $B_{bas}$  denote the *characteristic matrix* of the three basis points. From Eq (10) it follows that

$$B_{bas} = (P_{bas}^{-1})^T P_{bas}^{-1} \quad (30)$$

Finally, from the definition of pseudo-inverse it can be readily verified that

$$P^+ = (Q \cdot P_{bas})^+ = P_{bas}^{-1} Q^+ \quad (31)$$

We now describe  $B$  in terms of  $Q$  and  $B_{bas}$ . Substituting Eq (31) into the definition of  $B$  in Eq (10), and using Eq (30), we obtain

$$B = (P^+)^T P^+ = (Q^+)^T \cdot B_{bas} \cdot Q^+$$

The linear and incremental computation of the matrices  $Q$  and  $B_{bas}$  from at least three images of the object points is described in [Weinshall and Tomasi, 1992].

## C Eliminating translation

In this appendix we show that translation can be ignored if we set the centroids of both model and image points to be the origin. To show this, we prove that the best rigid and affine transformations maps the model centroid to the image centroid. We begin by showing that, given two sets of  $n$  2D points (images), the best translation that relates the two images maps the centroid of the first image to that of the second.

**Lemma 10:** Let  $p_1, \dots, p_n \in \mathcal{R}^2$  and  $q_1, \dots, q_n \in \mathcal{R}^2$  be two sets of corresponding points. Denote by  $\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i$  and  $\bar{q} = \frac{1}{n} \sum_{i=1}^n q_i$  the centroids of  $p_1, \dots, p_n$  and  $q_1, \dots, q_n$  respectively. The translation  $t^* \in \mathcal{R}^2$  that minimizes the term

$$D^* = \min_{t \in \mathcal{R}^2} \sum_{i=1}^n \|p_i + t - q_i\|^2$$

is given by

$$t^* = \bar{q} - \bar{p}$$

**Proof:** Assume, by way of contradiction, that the best translation is given by

$$t' = t^* + \delta$$

for some nonzero  $\delta \in \mathcal{R}^2$ . Denote the new term by  $D'$

$$\begin{aligned} D' &= \sum_{i=1}^n \|p_i + t' - q_i\|^2 \\ &= \sum_{i=1}^n \|p_i + t^* + \delta - q_i\|^2 \\ &= \sum_{i=1}^n \|p_i + t^* - q_i\|^2 + 2 \sum_{i=1}^n (p_i + t^* - q_i) \cdot \delta + \sum_{i=1}^n \|\delta\|^2 \\ &= D^* + 2n(\bar{p} + t^* - \bar{q}) \cdot \delta + n\|\delta\|^2 \end{aligned}$$

Since  $t^* = \bar{q} - \bar{p}$ , we obtain that

$$\bar{p} + t^* - \bar{q} = 0$$

and, therefore,

$$D' = D^* + n\|\delta\|^2$$

which implies that

$$D^* < D'$$

contradicting the initial assumption.

□

Using Lemma 10 we prove that the best rigid and affine transformations map the model centroid to the image centroid.

**Theorem 11:** Let  $P_1, \dots, P_n \in \mathcal{R}^3$  be a set of  $n$  model points, and let  $q_1, \dots, q_n \in \mathcal{R}^2$  be the corresponding  $n$  image points. The rigid transformation  $\{s^*, R^*, t^*\}$  that minimizes the term

$$D^* = \min_{\{s, R, t\}} \sum_{i=1}^n \|s\Pi R P_i + t - q_i\|^2$$

where  $\Pi$  denotes the orthographic projection, satisfies

$$\bar{q} = s^* \Pi R^* \bar{P} + t^*$$

**Proof:** Denote by  $p_i = s^* \Pi R^* P_i$ ; according to Lemma 10

$$t^* = \bar{q} - \bar{p}$$

Since

$$\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i = \frac{1}{n} \sum_{i=1}^n s^* \Pi R^* P_i = s^* \Pi R^* \bar{P}$$

we obtain that

$$\bar{q} = \bar{p} + t^* = s^* \Pi R^* \bar{P} + t^*$$

The theorem holds also if we consider affine transformations rather than only the rigid ones. The rotation matrix  $R$  is replaced in this case by a general linear transformation  $A$ .

□

Theorem 11 shows that the best rigid and affine transformations map the model centroid to the image centroid. Consequently, if the two centroids are moved to the origin, the translation component vanishes. This follows immediately from Theorem 11, since

$$\bar{q} = s^* \Pi R^* \bar{P} + t^*$$

then

$$\bar{P} = \bar{q} = 0$$

implies

$$t^* = 0$$

## D Best View

In this appendix we develop an expression for the **best view** of the transformation metric,  $N_{tr}$ . The derivations here follow the notations used in the proof of Theorem 1, from which we have that

$$\begin{aligned} s &= \frac{1}{2} \sqrt{w_1^2 + w_2^2 + 2w_1 w_2 \sin \theta} \\ s \cos \alpha &= \frac{1}{2} (w_1 + w_2 \sin \theta) \\ s \sin \alpha &= \frac{1}{2} w_2 \cos \theta \end{aligned}$$

According to Theorem 2,  $\bar{r}_1, \bar{r}_2 \in \text{span}\{\bar{a}_1, \bar{a}_2\}$ . We can therefore express  $\bar{r}_1$  and  $\bar{r}_2$  by

$$\bar{r}_1 = \beta_1 \bar{a}_1 + \beta_2 \bar{a}_2$$

$$\bar{r}_2 = \gamma_1 \bar{a}_1 + \gamma_2 \bar{a}_2$$

where  $\beta_1, \beta_2, \gamma_1$ , and  $\gamma_2$  are scalars. Substituting the definitions of the vectors  $\bar{r}_1, \bar{r}_2, \bar{a}_1$ , and  $\bar{a}_2$  we obtain

$$s \cos \alpha = \beta_1 w_1 + \beta_2 w_2 \cos \theta$$

$$-s \sin \alpha = \beta_2 w_2 \sin \theta$$

and

$$s \sin \alpha = \gamma_1 w_1 + \gamma_2 w_2 \cos \theta$$

$$s \cos \alpha = \gamma_2 w_2 \sin \theta$$

Therefore

$$\beta_1 = \frac{s \sin \alpha \cos \theta + s \cos \alpha \sin \theta}{w_1 \sin \theta}$$

$$\beta_2 = -\frac{s \sin \alpha}{w_2 \sin \theta}$$

$$\gamma_1 = \frac{s \sin \alpha \sin \theta - s \cos \alpha \cos \theta}{w_1 \sin \theta}$$

$$\gamma_2 = \frac{s \cos \alpha}{w_2 \sin \theta}$$

Substituting for  $s$  and  $\alpha$  we obtain

$$\beta_1 = \frac{1}{2} \left( 1 + \frac{w_2}{w_1 \sin \theta} \right)$$

$$\beta_2 = \gamma_1 = -\frac{\cos \theta}{2 \sin \theta}$$

$$\gamma_2 = \frac{1}{2} \left( 1 + \frac{w_1}{w_2 \sin \theta} \right)$$

And substituting for  $w_1, w_2$ , and  $\theta$

$$\beta_1 = \frac{1}{2} \left( 1 + \frac{\bar{y}^T B \bar{y}}{\sqrt{\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2}} \right)$$

$$\beta_2 = \gamma_1 = -\frac{\bar{x}^T B \bar{y}}{2 \sqrt{\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2}}$$

$$\gamma_2 = \frac{1}{2} \left( 1 + \frac{\bar{x}^T B \bar{x}}{\sqrt{\bar{x}^T B \bar{x} \cdot \bar{y}^T B \bar{y} - (\bar{x}^T B \bar{y})^2}} \right)$$



Now, to obtain the best view we use the following identities

$$\begin{aligned}\bar{x} &= P\bar{r}_1 & \bar{r}_1 &= \beta_1 \bar{a}_1 + \beta_2 \bar{a}_2 & \bar{a}_1 &= P^+ \bar{x} \\ \bar{y} &= P\bar{r}_2 & \bar{r}_2 &= \gamma_1 \bar{a}_1 + \gamma_2 \bar{a}_2 & \bar{a}_2 &= P^+ \bar{y}\end{aligned}$$

Therefore

$$\begin{aligned}\bar{x} &= PP^+(\beta_1 \bar{x} + \beta_2 \bar{y}) \\ \bar{y} &= PP^+(\gamma_1 \bar{x} + \gamma_2 \bar{y})\end{aligned}$$

## E Computing the eigenvalues of an ellipse

In this appendix we compute the eigenvalues of the ellipsoid  $B$  and the eigenvalue of an elliptic section of this ellipsoid.

We first show that the eigenvalues of the characteristic matrix,  $B$ , are  $\frac{1}{\lambda_1}$ ,  $\frac{1}{\lambda_2}$ , and  $\frac{1}{\lambda_3}$ , where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are the three positive eigenvalues of  $P^T P$ . This is derived as follows.

$$B\bar{a} = \frac{1}{\lambda} \bar{a} \iff P(P^T P)^{-1}(P^T P)^{-1}P^T \bar{a} = \frac{1}{\lambda} \bar{a}$$

Multiplying both sides by  $P^T$  we obtain that

$$(P^T P)^{-1}P^T \bar{a} = \frac{1}{\lambda} P^T \bar{a}$$

Denote  $\bar{b} = P^T \bar{a}$

$$(P^T P)^{-1} \bar{b} = \frac{1}{\lambda} \bar{b}$$

which implies that

$$(P^T P) \bar{b} = \lambda \bar{b}$$

Given  $\bar{X} = PP^+ \bar{x}$  and  $\bar{Y} = PP^+ \bar{y}$  in  $R^3$ , and a positive definite  $3 \times 3$  matrix  $B$ , let  $B'$  denote the ellipse defined by the intersection of the ellipsoid  $B$  with the plane  $\text{span}\{\bar{X}, \bar{Y}\}$ . We need to find the eigenvalues of  $B'$ ,  $\frac{1}{\mu_1}$  and  $\frac{1}{\mu_2}$ .

Without loss of generality we assume that  $\bar{X}$  and  $\bar{Y}$  lie on the ellipsoid defined by  $B$  (namely, we normalize the vectors so that  $\bar{X}^T B \bar{X} = \bar{x}^T B \bar{x} = 1$  and  $\bar{Y}^T B \bar{Y} = \bar{y}^T B \bar{y} = 1$ ). Let  $\theta$  denote the angle between  $\bar{X}$  and  $\bar{Y}$ . We define two orthonormal vectors  $\bar{x}'$  and  $\bar{y}'$ , which span the plane  $\text{span}\{\bar{X}, \bar{Y}\}$ , as follows:

$$\begin{aligned}\bar{x}' &= \frac{\bar{X}}{|\bar{X}|} \\ \bar{y}' &= \frac{\bar{Y} - \frac{\bar{X} \cdot \bar{Y}}{|\bar{X}|^2} \bar{X}}{|\bar{Y}| \sin \theta}\end{aligned}$$

Every vector  $\vec{v} \in \text{span}\{\vec{X}, \vec{Y}\}$  can be written as

$$\vec{v} = \alpha \vec{x}' + \beta \vec{y}'$$

and the intersection ellipse  $B'$  is given by

$$\vec{v} B' \vec{v} = 1 \iff (\alpha \ \beta) A^T B A \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 1$$

for  $A$  the  $3 \times 2$  matrix whose columns are  $\vec{x}'$  and  $\vec{y}'$ . We therefore have that

$$B' = A^T B A = \begin{pmatrix} (\vec{x}')^T B \vec{x}' & (\vec{x}')^T B \vec{y}' \\ (\vec{y}')^T B \vec{x}' & (\vec{y}')^T B \vec{y}' \end{pmatrix}$$

Substituting the expressions for  $\vec{x}'$  and  $\vec{y}'$ , we get

$$\begin{aligned} (\vec{x}')^T B \vec{x}' &= \frac{1}{|X|^2} \\ (\vec{y}')^T B \vec{y}' &= \frac{|X|^2 - 2|X||Y| \cos \theta (\vec{X}^T B \vec{Y}) + |Y|^2 \cos^2 \theta}{|X|^2 |Y|^2 \sin^2 \theta} \\ (\vec{x}')^T B \vec{y}' &= \frac{(\vec{X}^T B \vec{Y})|X| - |Y| \cos \theta}{|X|^2 |Y| \sin \theta} \end{aligned}$$

To obtain the two eigenvalues of  $B'$   $\frac{1}{\mu_1}$  and  $\frac{1}{\mu_2}$ , we solve the characteristic equation of  $B'$ , whose roots are

$$\frac{|X|^2 + |Y|^2 - 2|X||Y| \cos \theta \cdot \kappa \pm \sqrt{(|X|^2 + |Y|^2 - 2|X||Y| \cos \theta \cdot \kappa)^2 - 4|X|^2 |Y|^2 \sin^2 \theta (1 - \kappa^2)}}{2|X|^2 |Y|^2 \sin^2 \theta}$$

for  $\kappa = \vec{X}^T B \vec{Y} = \vec{x}'^T B \vec{y}'$ ,  $|X| = |PP^+ \vec{x}|$ ,  $|Y| = |PP^+ \vec{y}|$ , and  $\cos \theta = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}||\vec{y}|}$ .

## Acknowledgments

We thank Larry Maloney for valuable discussions and technical help, and Eric Grimson, Yael Moses, and Tomasso Poggio for their comments to earlier drafts of this paper

## References

- [Basri and Ullman, 1988] Basri, R. and Ullman, S. (1988). The alignment of objects with smooth surfaces. In *Proceedings of the 2nd International Conference on Computer Vision*, pages 482-488, Tarpon Springs, FL. IEEE, Washington, DC.

- [DeMenthon and Davis, 1992] DeMenthon, D. F. and Davis, L. S. (1992). Model-based object pose in 25 lines of code. In *Proceedings of the 2nd European Conference on Computer Vision*, Santa Margherita Ligure, Italy. Springer-Verlag.
- [Fischler and Bolles, 1981] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381-395.
- [Forsyth et al., 1991] Forsyth, D., Mundy, J. L., Zisserman, A., Coelho, C., Heller, A., and Rothwell, C. (1991). Invariant descriptors for 3-D object recognition and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:971-991.
- [Haralick et al., 1991] Haralick, R. M., Lee, C., Ottenberg, K., and Nolle, M. (1991). Analysis and solutions of the three point perspective pose estimation problem. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 592-598, Urbana-Champaign, IL.
- [Horaud et al., 1989] Horaud, R., Conio, B., Lebouilleux, O., and Lacolle, B. (1989). An analytic solution for the perspective 4-point problem. *Computer Vision, Graphics, and Image Processing*, 47:33-44.
- [Horn, 1987] Horn, B. K. P. (1987). Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4:629-642.
- [Horn, 1991] Horn, B. K. P. (1991). Relative orientation revisited. *Journal of the Optical Society of America*, 8:1630-1638.
- [Huttenlocher and Ullman, 1987] Huttenlocher, D. P. and Ullman, S. (1987). Object recognition using alignment. In *Proceedings of the 1st International Conference on Computer Vision*, pages 102-111, London, England. IEEE, Washington, DC.
- [Lamdan et al., 1987] Lamdan, Y., Schwartz, J. T., and Wolfson, H. (1987). On recognition of 3-d objects from 2-d images. Robotics research Technical Report 122, Courant Institute of Math. Sciences, N.Y. University.
- [Lowe, 1985] Lowe, D. G. (1985). Three-dimensional object recognition from single two-dimensional images. Robotics research Technical Report 202, Courant Institute of Math. Sciences, N.Y. University.
- [Rives et al., 1981] Rives, P., Bouthemy, B., Prasada, B., and Dubois, E. (1981). Recovering the orientation and the position of a rigid body in space from a single view. Technical report, INRS-Telecommunications, Quebec, Canada.
- [Roberts, 1965] Roberts, L. G. (1965). Machine perception of three-dimensional solids. In et al, T., editor, *Optical and Electro-Optical Information Processing*. M.I.T. Press, Cambridge, MA.
- [Strang, 1976] Strang, G. (1976). *Linear algebra and its applications*. Harcourt Brace Jovanovich, Orlando, Florida.

- [Thompson and Mundy, 1987] Thompson, D. W. and Mundy, J. L. (1987). Three-dimensional model matching from an unconstrained viewpoint. In *Proceedings of IEEE Conference on Robotics and Automation*, pages 208-220, Raleigh, NC.
- [Tsai, 1987] Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. of Robotics and Automation*, RA-3(4):323-344.
- [Ullman, 1989] Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition*, 32:193-254.
- [Ullman and Basri, 1991] Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:992-1006.
- [Weinshall, 1993] Weinshall, D. (1993). Model-based invariants for 3D vision. *International Journal on Computer Vision*, to appear.
- [Weinshall and Tomasi, 1992] Weinshall, D. and Tomasi, C. (1992). Linear and incremental acquisition of invariant shape models from image sequences. Research Report RC 18549, IBM T. J. Watson Research Center.
- [Yuan, 1989] Yuan, J. S. C. (1989). A general photogrammetric method for determining object position and orientation. *IEEE Trans. on Robotics and Automation*, 5(2):129-142.